3. Databases and Information Systems
An Improved Join Dependency For Efficient Constraint Checking

K. P. Tan  T. C. Tan

Department of Information Systems and Computer Science
National University of Singapore
Lower Kent Ridge, Singapore 0511, Singapore

Abstract: In a relational database model, checking the contraints of join dependency involves examining a set of n tuples and solving m1 constraint equalities. We derive a scheme called the \((n, m)\)-JD in which the number of constraint equalities is reduced to \(m^2\), by forming cyclic combinations of the (disjoint) elements of the partition and increasing the number of intersection operations to obtain the projections in the constraint equalities. The reduced set of constraint equalities results in less elementary checking operations and hence an overall increase in efficiency in the normalization effort. The relationship among \((n, m)\)-JDs of various degrees and orders is also studied and formalised.

Keywords: Database design, join dependency, constraint checking.

1. Introduction

The design of a database schema which does not contain any form of anomalies is a central problem in relational database design. Normalization theory, together with the technique of non-loss decomposition, has been devised as an aid to database design. Various normal forms have been introduced [Codd71] [Codd72] [Pagl77] [Nico78] [Dello78]. Rissanen introduced a data constraint known as the join dependency (JD) [Riss78], whose characteristics of lossless join decomposition leads to the formation of the fifth normal form [Pagl79]. Because of its importance, many papers have explored the properties and the complete axiomization of the JD, including [AhBU79], [BeVaS1], [Gyss85] and [Scio82]. Checking whether a relation instance is in the fifth normal form requires checking it against the JD. This paper extends the modified JD concept proposed in [Tan84] and also gives the general expression which relates the number of projections involved in a JD to the number of disjoint sets of attributes in each cyclic component.

2. Notations Used

Let \(U\), the universe, be a finite set of attributes. In this paper, we denote attributes by small letters, and sets of attributes by capital letters. So \(U = \{a, b, c, \ldots\}\). Also, we shall write AB for \(A \cup B\), where \(A\) and \(B\) are sets of attributes, and \(abc\ldots\) for \(\{a, b, c, \ldots\}\), where \(a, b, c, \ldots\) are attributes. Let \(R\) be the set of relation schemes over \(U\), and \(X \subseteq U\). A tuple \(t\) over \(X\) is a mapping that associates with each attribute \(a\) of \(X\) a value of its corresponding domain. An instance \(r\) of \(R\) over \(X\), denoted by \(r(X)\), is then a set of tuples over \(X\). Let \(Y \subseteq X\). The projection of a tuple \(t\) onto \(Y\), denoted by \(t[Y]\), is obtained by restricting the attributes of \(t\) to \(Y\). If \(r\) is an instance over \(X\), the set obtained by projecting each tuple of \(r\) onto \(Y\) is said to be the projection of \(r\) onto \(Y\), denoted \(\pi_Y(r)\).

The join dependency is defined as follows [Male83] [Ullm82]:

Let \(R = X_1, X_2, \ldots, X_n\) be a set of \(n\) relation schemes over \(U\), where \(X_1, X_2, \ldots, X_n \subseteq U\). A relation \(r(U)\) satisfies the \(n\)-JD, \(*[X_1, X_2, \cdots, X_n]*\) if \(r\) decomposes losslessly onto \(n\) projections \(X_1, X_2, \ldots, X_n : r = \pi_{X_1}(r) \otimes \pi_{X_2}(r) \otimes \cdots \otimes \pi_{X_n}(r)\). That is, \(r\) is the natural join of its projections onto the \(X_i\)'s.

In other words, if \(r\) contains tuples \(t_1, t_2, \ldots, t_n\) such that \(t_i(X_i \cap X_j) = t_j(X_i \cap X_j)\) for \(1 \leq i, j \leq n\), then \(r\) must contain a tuple \(t\) such that \(t(X_i) = t_i(X_i)\) for \(1 \leq i \leq n\).

We also define the operation \(\ominus\) as follows: for integers \(i, i' \leq n\), \(i \ominus i' = \begin{cases} i + i' & \text{if } i + i' \leq n \\ i + i' - n & \text{otherwise} \end{cases}\). \(\ominus\) is used in place of \(\ominus_n\) when there is no ambiguity.

3. \((n,m)\)-JD

We extend the discussion in [Tan84] to allow for the general case in which the cyclic component may have \((n-1)\) or less disjoint attribute sets, and compare the relationship between these variants. A new parameter \(m\) is installed in the new scheme for which we adopt the name \((n,m)\)-JD. The \((n,m)\)-JD is defined as follows:

Suppose \(R(U)\) is a relation over \(X_1, X_2, \ldots, X_n\) in the universe \(U\), where \(X_1, X_2, \ldots, X_n \subseteq U\), and \(U = \bigcup_{k=1}^{n} X_k\). Let \(U\) be partitioned into \(n\) disjoint sets \(Y_1, Y_2, \ldots, Y_n\), where \(U = \bigcup_{k=1}^{n} Y_k\) and \(Y_j \cap Y_k = \emptyset\) for all \(j \neq k\). In the sequel, we shall call \(n\),...